Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks

Lafci, Berkan ; Mercep, Elena ; Morscher, Stefan ; Deán-Ben, Xosé Luís ; Razansky, Daniel

Abstract: Multispectral optoacoustic tomography (MSOT) offers the unique capability to map the distribution of spectrally distinctive endogenous and exogenous substances in heterogeneous biological tissues by exciting the sample at various wavelengths and detecting the optoacoustically-induced ultrasound waves. This powerful functional and molecular imaging capability can greatly benefit from hybridization with pulse-echo ultrasound (US), which provides additional information on tissue anatomy and blood flow. However, speed of sound variations and acoustic mismatches in the imaged object generally lead to errors in the coregistration of compounded images and loss of spatial resolution in both imaging modalities. The spatially- and wavelength-dependent light fluence attenuation further limits the quantitative capabilities of MSOT. Proper segmentation of different regions and assignment of corresponding acoustic and optical properties turns then essential for maximizing the performance of hybrid optoacoustic and ultrasound (OPUS) imaging. Particularly, accurate segmentation of the boundary of the sample can significantly improve the images rendered. Herein, we propose an automatic segmentation method based on a convolutional neural network (CNN) for segmenting the mouse boundary in a pre-clinical OPUS system. The experimental performance of the method, as characterized with the Dice coefficient metric between the network output and the ground truth (manually segmented) images, is shown to be superior than that of a state-of-the-art active contour segmentation method in a series of two-dimensional (cross-sectional) OPUS images of the mouse brain, liver and kidney regions.

DOI: https://doi.org/10.1117/12.2543970

Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks

Lafci, Berkan, Merćep, Elena, Morscher, Stefan, Deán-Ben, Xosé Luís, Razansky, Daniel

Berkan Lafci, Elena Merćep, Stefan Morscher, Xosé Luís Deán-Ben, Daniel Razansky, "Efficient segmentation of multi-modal optoacoustic and ultrasound images using convolutional neural networks," Proc. SPIE 11240, Photons Plus Ultrasound: Imaging and Sensing 2020, 112402N (17 February 2020); doi: 10.1117/12.2543970

SPIE.

Event: SPIE BiOS, 2020, San Francisco, California, United States

Efficient Segmentation of Multi-modal Optoacoustic and Ultrasound Images Using Convolutional Neural Networks

Berkan Lafci[a,b], Elena Merćep[c], Stefan Morscher[c], Xosé Luís Deán-Ben[a,b], and Daniel Razansky[a,b,*]

[a]Institute for Biomedical Engineering and Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland
[b]Faculty of Medicine and Institute of Pharmacology and Toxicology, University of Zurich, Switzerland
[c]iThera Medical GmbH, Munich, Germany

ABSTRACT

Multispectral optoacoustic tomography (MSOT) offers the unique capability to map the distribution of spectrally distinctive endogenous and exogenous substances in heterogeneous biological tissues by exciting the sample at various wavelengths and detecting the optoacoustically-induced ultrasound waves. This powerful functional and molecular imaging capability can greatly benefit from hybridization with pulse-echo ultrasound (US), which provides additional information on tissue anatomy and blood flow. However, speed of sound variations and acoustic mismatches in the imaged object generally lead to errors in the coregistration of compounded images and loss of spatial resolution in both imaging modalities. The spatially- and wavelength-dependent light fluence attenuation further limits the quantitative capabilities of MSOT. Proper segmentation of different regions and assignment of corresponding acoustic and optical properties turns then essential for maximizing the performance of hybrid optoacoustic and ultrasound (OPUS) imaging. Particularly, accurate segmentation of the boundary of the sample can significantly improve the images rendered. Herein, we propose an automatic segmentation method based on a convolutional neural network (CNN) for segmenting the mouse boundary in a pre-clinical OPUS system. The experimental performance of the method, as characterized with the Dice coefficient metric between the network output and the ground truth (manually segmented) images, is shown to be superior than that of a state-of-the-art active contour segmentation method in a series of two-dimensional (cross-sectional) OPUS images of the mouse brain, liver and kidney regions.

Keywords: Optoacoustic imaging, Ultrasound Imaging, Concave arrays, Deep Learning, Segmentation

1. INTRODUCTION

Multispectral optoacoustic tomography (MSOT) can differentiate tissue characteristics and quantify functional parameters by capitalizing on the absorption properties of endogenous and exogenous substances for various excitation wavelengths. MSOT has then been successfully applied on various clinical and preclinical studies that require molecular and functional information to quantify disease biomarkers.[1–4] MSOT additionally provides anatomical information related to the blood vessel distribution, which may however be insufficient for the identification of certain organs and structures.[5] Anatomical images acquired with pulse-echo ultrasound can complement MSOT and aid to the interpretation of functional and molecular parameters. In addition, the complementary information rendered with both modalities facilitates improving reconstruction of optoacoustic (OA) and US images with more accurate acoustic inversion methods. These advantages have fostered the development of OPUS systems by several groups.[6–9] Image registration and reconstruction generally requires knowledge on the object boundaries. Particularly, accurate segmentation of outer boundary can be used to assign different

Further author information: (Send correspondence to D.R.)
B.L.: e-mail: blafci@student.ethz.ch, E.M.: elena.mercep@ithera-medical.com, S.M.: stefan.morscher@ithera-medical.com, X.L.D.B.: e-mail: xl.deanben@pharma.uzh.ch, D.R.: e-mail: daniel.razansky@uzh.ch

Photons Plus Ultrasound: Imaging and Sensing 2020, edited by Alexander A. Oraevsky, Lihong V. Wang,
Proc. of SPIE Vol. 11240, 112402N · © 2020 SPIE · CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2543970

values of acoustic and optical properties in the sample and the coupling medium (typically water), which has been shown to significantly improve the quality of the images.[10,11]

In this work, we suggest a convolutional neural network (CNN)-based segmentation approach to accurately find the object boundaries in OPUS images. The performance of the segmentation algorithm was analyzed quantitatively by considering the Dice coefficient.[12] The results obtained were also compared with those obtained with active contour, a widely-used image segmentation method of medical images. The active contour algorithm was initialized by edge detection to remove user dependency. Fully automated CNN-based segmentation algorithm is shown to have an improved performance compared to active contour with edge detection on both OA and US images.

2. METHODS

2.1 Convolutional Neural Network

CNNs apply convolutional image kernels to extract features from images. These kernels are learned during the training process and adapted to implement specific tasks, where training is performed on batches of images by looping through several epochs. The trained network is then able to perform the same task on other images. The organization of convolutional image kernels of different sizes and the combination of nonlinear units defines the CNN architecture. In this study, the so-called U-Net architecture was used.[13] This contains two parts termed encoder and decoder. The encoder part gets the input images and scales them down by increasing the channel number. In contrast, the decoder part increases the image dimensions and decreases the channel number until it gets the dimensions of the data fed into the network. The loss function between ground truth image and the prediction of the network is then calculated. Herein, a combination of binary cross entropy and soft Dice loss with equal weight was used for training the U-Net network. The loss function was optimized using stochastic gradient descent (SGD) with learning rate of 0.01 and momentum 0.99. The weights were randomly initialized and the network was trained with batch size of 5 over 100 epochs. PyTorch library was used for implementation.

The U-Net architecture was adjusted in terms of input image size. Specifically, 256x256 pixel images with channel number 1 were used as input to the network, which gives as output images of the same size. In the encoder part, convolutional layers of 3x3 pixel image kernels with stride 1 were applied followed by the 2x2 max pooling layers with stride 2. These layers were followed by batch normalization and ReLU layers.[14] In the decoder part, transposed convolution kernels with size 3x3 pixels and stride of 1 were applied followed by up-sampling using bilinear interpolation. The corresponding images from the encoder part were then concatenated with up-sampled images and convolved with the same size of kernels in the encoder part. These convolutional layers were followed by batch normalization and ReLU layers. The network contained four down-sampling blocks in encoder and four up-sampling blocks in decoder.

2.2 Active Contour with Edge Detection

The CNN-based segmentation method was compared with the active contour method with edge detection (ACED). The ACED algorithm was implemented as follows. It first finds the edges in a given image and fits a circle containing all detected edges. This circle is then given to the active contour algorithm implemented in Li et al.[15] By initializing the algorithm with automatic edge detection, we guaranteed that both CNN-based and ACED-based segmentation algorithms do not require user input and are fully automated.

The ACED algorithm received OA and US images with 256x256 pixels that are downscaled to 150x150 pixels in order to reduce computation power needed for calculation of several iterative steps. After downscaling, a Gaussian filter with kernel size 3x3 pixels and sigma 0.5 was applied on the imaging for smoothing. Then, Canny edge detector[16] was used to find structures identified as edges in the images associated to high intensity gradients. Morphological operations using dilation and erosion with disc shaped structuring element with size of 3x3 pixels were applied on the structures in order to remove outliers and non-connected edges. The initial circle encapsulating the remaining structures was fitted and used as input of the spline fitting algorithm. The spline contour evolves over 20 iterations by using 1-pixel neighborhood.

2.3 Dataset

The dataset contains OA and US cross-sectional images from the brain, kidney and liver regions of mice. All images were acquired with inVision 256 TF and inVision 512-echo MSOT and OPUS scanners (iThera Medical GmbH, Munich, Germany). All animal experiments were conducted under the institutional guidelines for the use of laboratory animals and approval from the government of upper Bavaria.

OA imaging was performed by tuning the excitation laser to 800 nm wavelength. The detected signals were band-pass filtered with low cut-off frequency 50 kHz and high cut-off frequency 6 MHz, and subsequently deconvolved with the impulse response of the transducer. The filtered signals were processed using a back-projection algorithm to generate images with 100 μm resolution and 30 mm^2 field of view (FOV). Pulse-echo US signals were reconstructed using the synthetic aperture (STA) method, which is based on transmitting waves from single element at a time and collecting US echoes with the consecutive elements around the transmitting element. The US signals were processed using conventional delay and sum algorithm for each transmission event, and the resulting images from every transmission event were compounded to create final pulse-echo US image.

OA and US images were separated into two different groups corresponding to training and testing. The separation was done based on the mouse ID to avoid including the same mice in the training and test sets. Only OA images were used for training the network, which was tested on both OA and US images. The number of the images and mice that were used for network training and test are summarized in Table 1.

Table 1. Number of the images that are used for the training and test of the network.

Anatomical Region	Number of training images	Number of test images	
	OA Mode	OA Mode	US Mode
Brain	174 (12 mice)	28 (4 mice)	9 (2 mice)
Kidney	97 (13 mice)	38 (5 mice)	16 (5 mice)
Liver	108 (15 mice)	33 (5 mice)	30 (5 mice)
Total	379	99	55

2.4 Quantitaive Evaluation

The segmentation performance was evaluated based on the Dice coefficient metric, which uses the ratio of true positive (TP), false positive (FP) and false negative (FN) pixels. The TP, FP and FN pixels were determined by comparing the manually segmented ground truth data with the output of the segmentation algorithms. The Dice coefficient is defined as follows

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}.$$

3. RESULTS

The segmentation results of ACED and CNN-based methods on OA images are summarized in Fig. 1. The left column (Figs. 1a, 1d, 1g) shows the ground truth data corresponding to the manually segmented images used for the quantitative evaluation of the results. The middle column (Figs. 1b, 1e, 1h) displays the results obtained with the ACED segmentation method. As shown in Fig. 1b, ACED can converge to a smaller region than that defined with the actual boundary. On the other hand, Fig. 1e and Fig. 1h show that it can also result in bigger or shifted regions compared to ground truth data. The respective Dice coefficients for Figs. 1b, 1e and 1h are 0.72, 0.83 and 0.78 respectively. CNN based segmentation outperforms ACED segmentation in all cross sections. CNN predicts similar regions to those obtained with manual segmention in Figs. 1c, 1f and 1i. The Dice coefficients on cross sections of brain, liver and kidney are 0.98, 0.96 and 0.97, respectively.

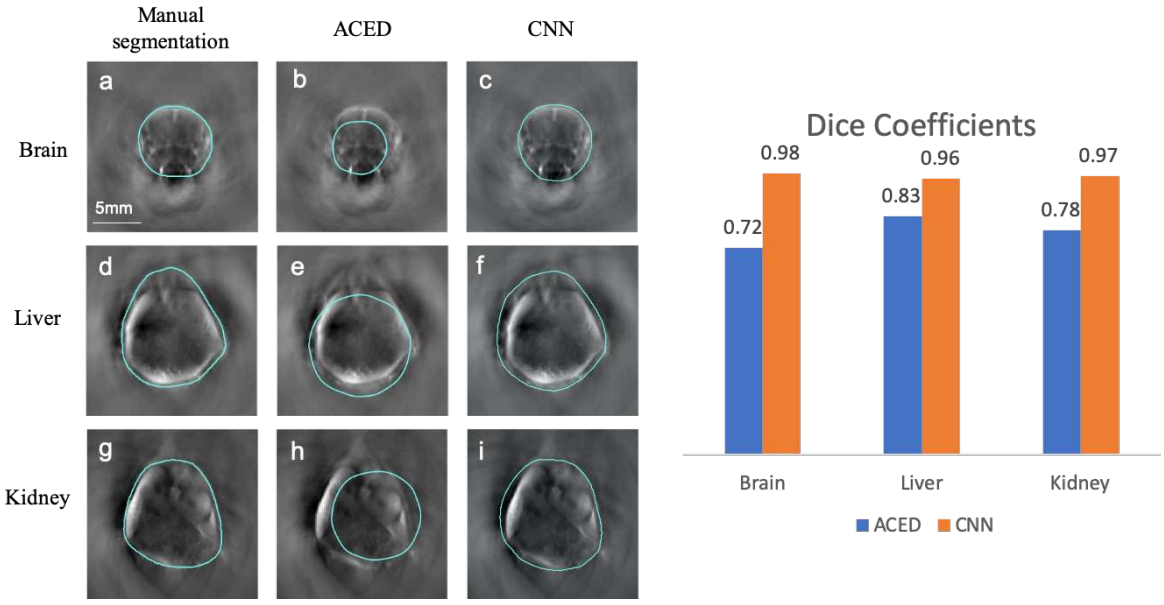

Figure 1. Results of segmentation algorithms on the optoacoustic images.

The results of both segmentation algorithms on US images are shown in Fig. 2. The manually segmented boundaries are shown in Figs 2a, 2d and 2g. It is shown that the ACED algorithm fits larger circles than the expected input, arguably because of the image artifacts around the mice boundaries. This larger circle used for initialization results in all the pixels being included into the segmentation mask. The values of Dice coefficients are 0.34, 0.61 and 0.59 for brain, liver and kidney cross sections, respectively. The CNN-based segmentation algorithm outperforms the ACED algorithm on US images, but the network gives less accurate results on US images compared to OA test images. The main reason behind this is probably related to the fact that the network does not see any images with similar image intensity and noise distribution during the training process. The network trained with OA images resulted in Dice coefficients of 0.70, 0.82 and 0.86 on US brain, liver and kidney US images respectively.

The overall performances of CNN and ACED algorithms on all test images was also compared by calculating the corresponding Dice coefficients. The CNN based segmentation algorithm achieved Dice coefficients of 0.95±0.04 on OA test images, where values of 0.78±0.11 were obtained with the ACED method. On the other hand, the ACED algorithm achieved Dice coefficients of 0.55±0.10 on US images versus values of 0.80±0.08 obtained with CNN-based image segmentation.

4. CONCLUSIONS

The application of CNNs for the segmentation of images rendered with different modalities showed promising results in the medical imaging field.[17–19] Herein, we applied CNNs for segmenting object boundaries in OA and US images. We have shown improved performance of CNNs on OA and US images compared to the widely-used ACED method. Specifically, CNN was shown to render more accurately segmentation and less outliers in cross sections of the mouse corresponding to the brain, liver and kidney regions. In contrast, the ACED algorithm showed less accurate segmentation of mouse boundaries and rendered more artifacts in the images. In this work, CNN was trained by considering only OA images because of the limited number of US images. Relatively accurate US image segmentation was still obtained under these circumstances. Higher accuracy on US image segmentation may be achieved by training the network with a combination of OA and US images or separately with a set of US images. Also, the performance of the CNN algorithm may be improved with more OA images in training set.

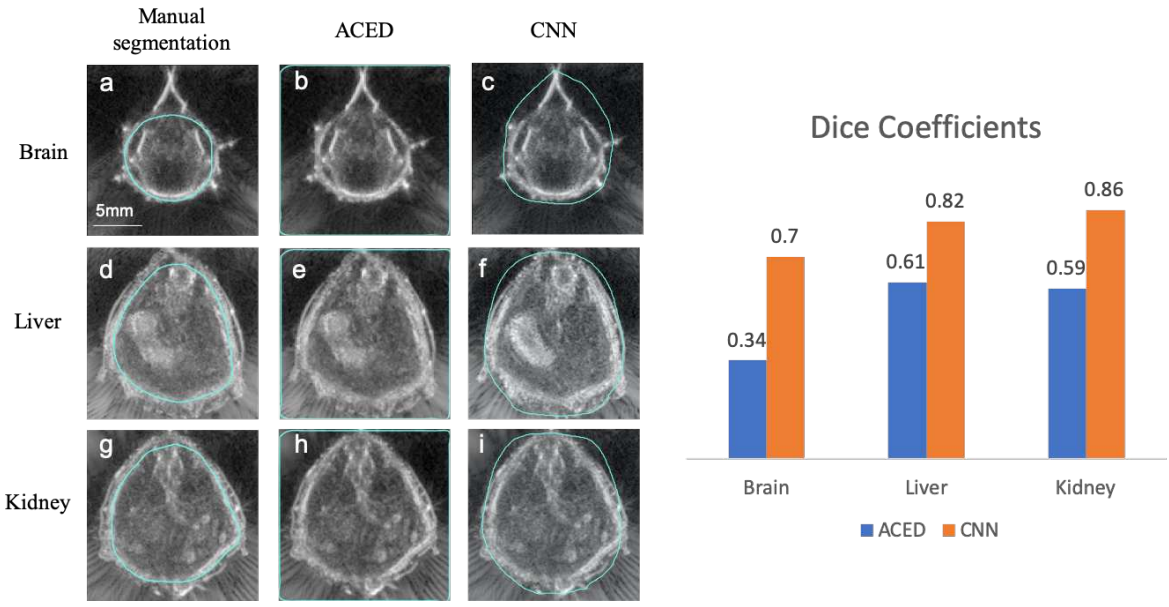

Figure 2. Results of segmentation algorithms on the ultrasound images.

REFERENCES

[1] Tomaszewski, M. R., Gonzalez, I. Q., O'Connor, J. P., Abeyakoon, O., Parker, G. J., Williams, K. J., Gilbert, F. J., and Bohndiek, S. E., "Oxygen enhanced optoacoustic tomography (oe-ot) reveals vascular dynamics in murine models of prostate cancer," *Theranostics* **7**, 2900–2913 (07 2017).

[2] Yao, J. and Wang, L. V., "Recent progress in photoacoustic molecular imaging," *Current opinion in chemical biology* **45**, 104–112 (08 2018).

[3] Ron, A., Deán-Ben, X. L., Gottschalk, S., and Razansky, D., "Volumetric optoacoustic imaging unveils high-resolution patterns of acute and cyclic hypoxia in a murine model of breast cancer," *Cancer Research* **79**(18), 4767–4775 (2019).

[4] Ivankovic, I., Deán-Ben, X. L., Lin, H.-C. A., Zhang, Z., Trautz, B., Petry, A., Görlach, A., and Razansky, D., "Volumetric optoacoustic tomography enables non-invasive in vivo characterization of impaired heart function in hypoxic conditions," *Scientific reports* **9**, 8369–8369 (06 2019).

[5] Merčep, E., Jeng, G., Morscher, S., Li, P. C., and Razansky, D., "Hybrid optoacoustic tomography and pulse-echo ultrasonography using concave arrays," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **62**, 1651–1661 (Sept 2015).

[6] Choi, W., Park, E.-Y., Jeon, S., and Kim, C., "Clinical photoacoustic imaging platforms," *Biomedical engineering letters* **8**, 139–155 (04 2018).

[7] Nuster, R., Schmitner, N., Wurzinger, G., Gratt, S., Salvenmoser, W., Meyer, D., and Paltauf, G., "Hybrid photoacoustic and ultrasound section imaging with optical ultrasound detection," *Journal of Biophotonics* **6**(67), 549–559 (2013).

[8] Deán-Ben, X. L., Merčep, E., and Razansky, D., "Hybrid-array-based optoacoustic and ultrasound (opus) imaging of biological tissues," (2017).

[9] Merčep, E., Herraiz, J. L., Deán-Ben, X., and Razansky, D., "Transmission-reflection optoacoustic ultrasound (tropus) computed tomography of small animals," *Light, science & applications* **8**, 18–18 (01 2019).

[10] Brochu, F. M., Brunker, J., Joseph, J., Tomaszewski, M. R., Morscher, S., and Bohndiek, S. E., "Towards quantitative evaluation of tissue absorption coefficients using light fluence correction in optoacoustic tomography," *IEEE Transactions on Medical Imaging* **36**, 322–331 (Jan 2017).

[11] Mandal, S., Den-Ben, X. L., and Razansky, D., "Visual quality enhancement in optoacoustic tomography using active contour segmentation priors," *IEEE Transactions on Medical Imaging* **35**, 2209–2217 (Oct 2016).

[12] Dice, L. R., "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).

[13] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention (MICCAI)*], *LNCS* **9351**, 234–241, Springer (2015).

[14] Nair, V. and Hinton, G. E., "Rectified linear units improve restricted boltzmann machines," in [*Proceedings of the 27th International Conference on International Conference on Machine Learning*], *ICML10*, 807–814, Omnipress, Madison, WI, USA (2010).

[15] Li, C., Xu, C., Gui, C., and Fox, M. D., "Distance regularized level set evolution and its application to image segmentation," *IEEE Transactions on Image Processing* **19**, 3243–3254 (Dec 2010).

[16] Canny, J., "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, 679–698 (Nov 1986).

[17] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J., "Deep learning for brain mri segmentation: State of the art and future directions," *Journal of Digital Imaging* **30**, 449–459 (Aug 2017).

[18] Blanc-Durand, P., Gucht, A. V. D., Schaefer, N., Itti, E., and Prior, J. O., "Automatic lesion detection and segmentation of 18f-fet pet in gliomas: A full 3d u-net convolutional neural network study," *PLOS ONE* **13**, 1–11 (04 2018).

[19] Christ, P. F., Ettlinger, F., Grün, F., Elshaer, M. E. A., Lipková, J., Schlecht, S., Ahmaddy, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Hofmann, F., D'Anastasi, M., Ahmadi, S. A., Kaissis, G., Holch, J. W., Sommer, W. H., Braren, R., Heinemann, V., and Menze, B. H., "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," *CoRR* **abs/1702.05970** (2017).